MEDICAL DEPARTMENT

U.S. NAVAL AIR STATION

QUONSET POINT, R. I.

16 April 1945

THE EFFECT OF NIGHT VISION TRAINING (EVELYN TRAINER)

ON U..S. NAVY RADIUM PLAQUE ADAPTOMETER SCORES.

Research Project: BuM&S No. X-558 (W-289-p)

Report No. One

Principal Investigator: J. Orlansky, Lt.(jg) H(S), USNR

Collaborator: W. Dana, Captain, (MC), USN

APPROVED:

  W. DANA
  Captain (MC) USN

WW
145
gu58e
1945

Film 958, Item 8

JUL 10 1950

U.S. NAVAL AIR STATION

QUONSET POINT, R.I.

OBJECT: To investigate the effect of the Evelyn two-dimensional night
            vision trainer on scores on the Navy Radium Plaque Adaptometer
            (RPA) Test.

METHOD: To compare with a control group the percent failing the RPA
            Test after training.

SUMMARY AND
CONCLUSIONS:

Following one session on the Evelyn two-dimensional night
vision trainer, a group exhibited 16.6% failures compared
to 18.4% failures in a control group on the RPA test.

When the test immediately follows training, there are less
failures (14.9%) than when there is a delay of one week
(19.6%).

Men who have previously passed the RPA, are then Evelyn trained
and subsequently RPA retested still show 8.3% failures.

In a small group, 13.8% of previous passers fail the RPA
upon retest. None of these men were Evelyn trained.

While few of these findings are statistically reliable, the
observed tendencies are consistent and in the same direction
when measured by several indicators. The same tendencies are
observed even after those individuals getting a score of 0/20
are arbitrarily excluded.

452316

CONFIDENTIAL

BACKGROUND: This study was designed to discover whether the number of men
    ordinarily failing the Navy Radium Plaque Adaptometer Test (Device 9-B-4)
    might be reduced if this test were preceded by training on the Evelyn
    Night Vision Trainer (Device 9-W). If training improves group performance
    on the night vision test, one might then prescribe training to limit the
    number of failures. While the incidence of failures could be limited
    arbitrarily by brightening the Plaque or by liberalizing the performance
    requirements on the present test, training might insure that personel
    approach the test with a more equal and relevant background of experience.
    Should this be confirmed by evidence, the test would then intrinsically
    be more uniform for all and performance thereon a more significant indi-
    cation of night vision ability.

EXPERIMENTAL PROCEDURE: During the course of this experiment, men who appeared
    for night vision training were required also to be tested for their night
    visual ability. In some cases the test immediatly followed the training,
    while in others the test came after a week's delay. This permitted an
    estimate not only of the effect of training but of the duration of such
    an effect as well. All subjects were questioned concerning previous night
    vision training and testing they might have encountered. As a result of
    this information, some observations are possible concerning the effect of
    a previous RPA test on subsequent RPA scores. None of the subjects had en-
    countered any form of night vision training prior to this experiment.

    While it might be of great interest to measure the effectiveness of the
    Evelyn trainer on night visual performance--and this is partly done in
    the present study--the principal aim was to ascertain whether the RPA
    test results might be influenced by some training procedure. Accordingly,
    the percent normally failing the RPA may be taken as the standard against
    which comparisons may be evaluated. Information based on the normal test-
    ing of 4477 men at this Station was available and used for this purpose.

    Both the Evelyn training and the RPA test were performed routinely on the
    men involved in this study. Adequate dark adaptation was insured by having
    the subjects wear red goggles for twenty minutes and then spend ten more
    minutes in complete darkness before being tested. The Evelyn training ses-
    sion, which lasted about an hour, demonstrated the need for dark adaptation,
    peripheral vision and scanning procedures.

    In all instances, the Evelyn training proceeded the night vision test and,
    indeed, some men were led directly from one to the other. These constitute
    our "immediately tested" group. Some other men were tested a week after
    the training session.

    Among our subjects were some men for whom this experiment meant that they
    were taking the RPA test for the second time and these have been treated
    separately in the data.

- -

RESULTS: The principal findings are summarized in Table 1. A total of 648 men were tested.

The performance of these subjects may be assessed by comparing them with the records of groups that have been routinely run through the RPA at this Station. During the period of June 1944 to February 1945, 4477 men were tested and, of these, 18.4% failed to reach the standard Navy performance. (Fig 1) The median score for the whole group is 18.28 correct out of 20. This figure approximates that reported by the U.S. Submarine Base at New London, Connecticut, where 17.6% men fail in a sample of 5750 (1). The present authors have collected evidence which suggests that the number of men failing the RPA test varies with the seasons of the year, the significant factor probably being exposure to sunlight; this will be the subject of a separate report. The matter is pertinent to this report in setting up a proper control against which the experimental effects may be estimated. A sample of 1297 men were tested here during January and February 1945, while out data were being collected, shows that 18.7% failed. As one would expect, this figure is not significantly different from the average for the whole group; (the chi-square test yields p = .77).

One experimental group of 434 men showed 16.6% failures on the RPA after having been trained, instead of the normal 18.4%. Its median score was 18.61 and this is somewhat higher, i.e., better, than 18.28 in the control group. (Fig 2) This information becomes more meaningful if analyzed to reveal the influence of a time delay between training and the test. When the test immediately follows the training session, as it did for 276 men, there are 14.9% failures and a median of 18.69. After a delay of one week for 158 men, the failures rise to 19.6% and the median drops to 18.44. None of the percent deviations from the control measures are statistically significant by the chi-square test, though the difference achieved by the immediately tested group approaches this with a "p" equal to .15. Such an indication may be taken to mean that if there are any effects of training as can be measured by the RPA, these are short-lived indeed.

By far, the best performance was turned in by a group of 134 men, who, prior to this experiment, had once passed the RPA test. When these men were tested again, in this case immediately after the Evelyn training, only 8.3% failed and the median score is 19.03; chi-square shows the difference to be reliable with a "p" equal to .01. This drop in the number of failures may be due either to the RPA retest as such, to the selection of previous passers, and to the Evelyn training. A suggestion as to the relative involvement of these factors may be gained by inspecting our final group of 80 men, all previous RPA test "passes", who were retested though not Evelyn trained. Here there are 13.8% failures, a drop from the 18.4% normal group as substantial in magnitude as that due to training (14.9% for one sub-group, 16.6% for the entire group). With both groups identical in all respects except for the matter of Evelyn training, the difference between 13.8% and 8.3% failures may well be an indication of the influence of Evelyn training; while the difference between 18.4% for the normal group and 13.8% for the retest group may

be partly due to the retest. By far the more disturbing fact, however, is that 13.8% of a group that once passed the RPA may fail on retest.

As reported above, the variations in both the percent failing and the median score may be generally interpreted to point to the same tendency. To a lesser extent, this is true as well for other indications of performance that may be found in the percent achieving a perfect score (10/10) and in the percent failing to see anything at all (0/20). (Fig. 3) Cumulative percent curves in Fig. 4 also show the separation between the several groups.

DISCUSSION:

Since few of the observed differences in percent failing are reliably differentiated from the control measures, some caution must be exercised in interpreting these results. Not only are the samples relatively small, but the RPA, as an experimental indicator, does not mirror all the functions one is likely to train in night vision. One night also prefer the RPA to have a higher test-retest reliability. However, there are certain clear tendencies in the data, and another investigation with larger groups might yield not only direct verification but even statistical reliability for differences of the same magnitude.

with these cautions in mind, one may assert that Evelyn training only slightly (1.8%) diminishes the incidence of failing the RPA, though the effect is larger (3.5%) when the test immediately follows training. This finding is a disappointment should one propose use of the Evelyn to reduce sharply the number of failures. A broader view may be that the substance of Evelyn training can improve the lower threshold of night vision, one may assume that the effective carry-over to the test is some knowledge about peripheral fixation and possibly a greater motivation to succeed. A single session in the two-dimensional trainer scarcely affords much practice in scanning or peripheral fixation, and additional training may be necessary to instill these as working habits. As measured on the RPA, the effect of training wears off rapidly, which fits in with these observations. Training more likely to decrease RPA failures would have to provide experience relevant to the test, and this information already exists. Thus, the New London Sub Base reports that when failures are retested, only 5% of the population still fail after three tests. (1)

It may well be that the performance required on the RPA is so simple and basic that, except for the matter of fixation, it is only relatively susceptible to improvement by training. For this purpose, direct practice in fixation, repeating the test and greater emphasis on fixation by the test operators would result in a more uniform performance. By the same fact, however, one should not expect it to give the best all-around prediction of night look out ability.

The question as to just what the two and three dimensional trainers achieve except for demonstration, remains unanswered and is not properly the concern

of this study. However, such an investigation is imperative for testing
the present effectiveness of the devices, for developing an inclusive
curriculum, for establishing practice procedures, and for instituting a
level of competence which personnel must reach before being qualified for
night lookout duties.

Indeed, the least failures (8.3%) are among those who, having previously
passed the RPA, have now been trained and retested. Selection, training
and retest all contribute to the reduction, the intermingling of which may
be partly judged by the percent of passers who fail on retest alone (13.8%)
without benefit of training. One possible flaw in these data is that it was
impossible to check in all cases the subjects' claim of a previous passing
performance on the RPA. Let us note at least, that the group one would ex-
pect to be most favored, with selection, training and retest, still exhibits
8.3% failures. Evelyn training an unselected group could hardly reduce
beyond that figure the percentage of failures on the first RPA.

On a purely statistical basis the evaluation of these results must be
tempered with caution. Five out of twenty-one different possible scores
constitute a passing performance, while the remaining sixteen are called
failures. The resulting frequency distribution is hardly normal and this
is a function, mainly, of the failure to discriminate effectively among those
who pass (approximately 82% of the population). To a lesser extent, another
contribution to distortion of the curve is made by a segment of the population
(approximately 8%) who probably feign night blindness and get scores of 0/20.
This casts doubt over any statistical procedure based on the expectation of
a normal distribution.

Since none of the observed effects in this study may be said to be unduly
large, the existence of a large group of individuals who fail in the 0/20
category, i.e., by failing to guess right on the test not once in 20 trials,
constitutes a hazard for any interpretation. On a chance basis, there are
three ways out of four than an individual may be wrong on any single trial
and 3 chances in 1000 that this could occur on 20 straight trials. Some-
thing other than chance must be operating when 8% rather than an expected
.3% of the group get a score of 0/20. ...en examined on the RPA may feel that
a failure on the test will save them from going to sea and that there is an
advantage in not being able to see anything at all. The method for achieving
this is naive when the individual for 20 trials calls the letter "T" in a
position directly opposite from that in which it is placed. Charting the
dark adaptation curve or scotopic perimetry might detect the fraudulent. The
presence, however, of such a group in an experiment might easily mean that
the observed results are due less to the influences being studied than to the
capricious nature of these examinees. In the 9 months during which the RPA
has been employed at this station, the number of men getting this score has
varied from 0 to 12.0%, with a mean of 8.5% for the total group.

we assumed, therefore, that most of these men were probably feigning night
blindness and desired to evaluate the data in a way to be free of any in-
fluence these men might contribute. A limit of 2% failing completely (0/20)
was established, and those making up more than that percent of the population

-5-

in each group arbitrarily excluded from the data, which was then reworked
and presented in Table 2. The main finding is that the tendencies originally
observed still remain and we may, with some confidence, assert that the data
is not an artifact due entirely to the chance composition of the 0/20
groups. (Fig. 2, 3).

REFERENCES:

    (1) A report on the night vision testing of 5750 men, Medical Research
        Dept., U.S. Submarine Base, New London, Connecticut, 12 Aug 1944.

TABLE I

Summary of Data

	N	% Fail	Median Score	% 10/10 scores	% 0/20 scores	P value x^2 test for against control
Control - Quonset Pt.	4477	18.4	18.28	41.4	8.5	
Evelyn trained, RPA tested whole group	434	16.6	18.61	45.2	7.6	.35
Evelyn trained, tested immediately	276	14.9	18.69	46.0	8.0	.15
Evelyn trained tested one week later	158	19.6	18.44	43.6	7.0	.65
Previous RPA pass, Evelyn trained, RPA retest	134	8.3	19.03	51.5	5.2	.01
No Evelyn training, RPA test and retest	80	13.8	19.11	56.3	7.5	.25

* when this value reaches .05 or less, then differences as large as this due to chance factors could have arisen only 5 times in 100 and the difference may then be considered reliable.

TABLE II

Summary of Data

Recalculated after limiting failures
in the 0/20 category to 2% of the population

	N	% Fail	Median Score	% 10/10	% 0/20	P₂value χ^2 test against control
Control – Quonset Pt.	4188	11.9	18.55	44.2	2.0	
Evelyn trained, RPA Tested, whole group	410	11.7	18.83	47.8	2.0	.88
Evelyn trained, tested immediately	260	9.6	18.92	48.9	2.0	.25
Evelyn trained, tested one week later	150	15.3	18.67	46.0	2.0	.20
Previous RPA pass, Evelyn trained, RPA retest	130	5.4	19.06	53.1	2.0	.03
No Evelyn training, RPA test and retest	78	9.2	19.16	57.7	2.0	1.00

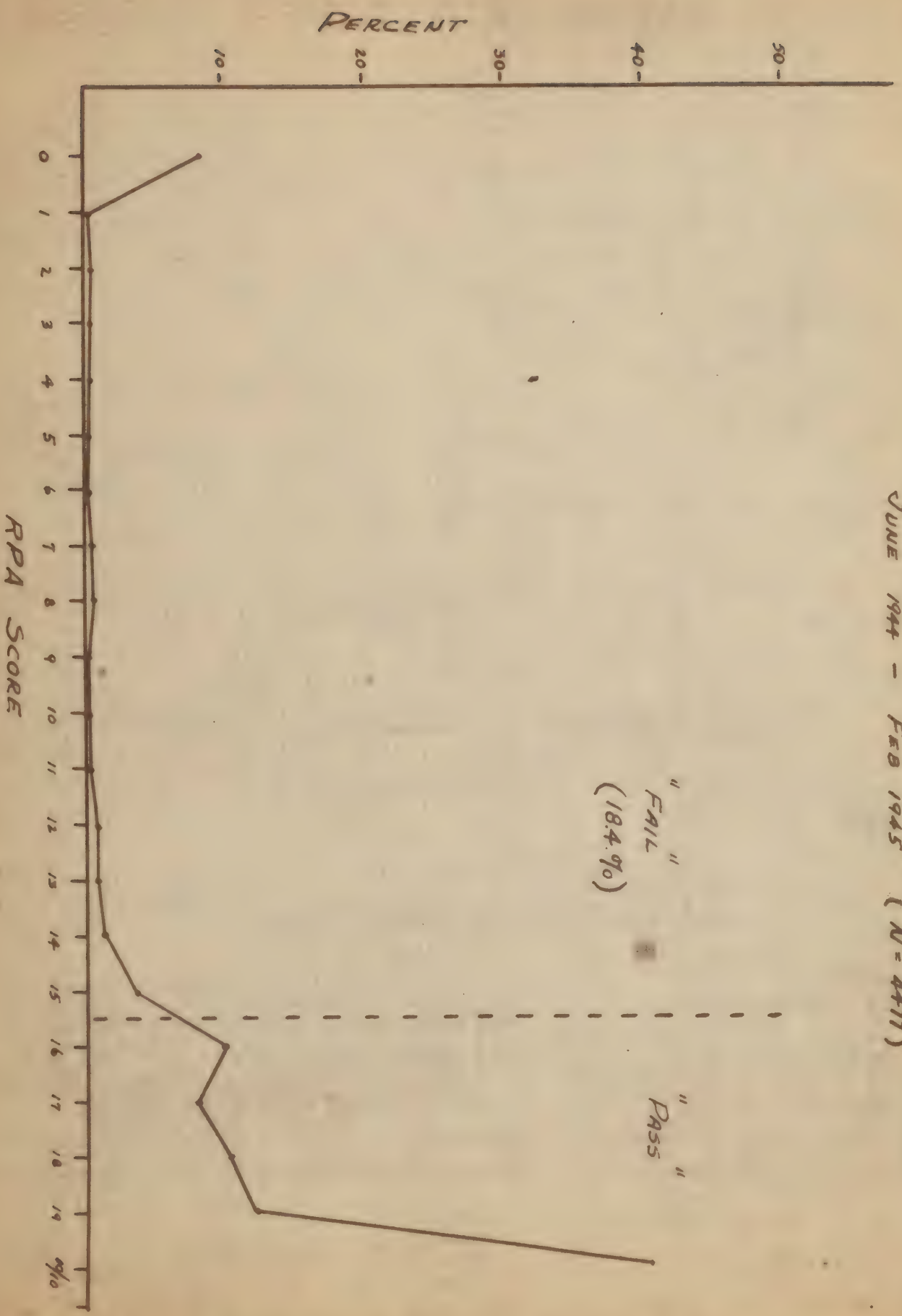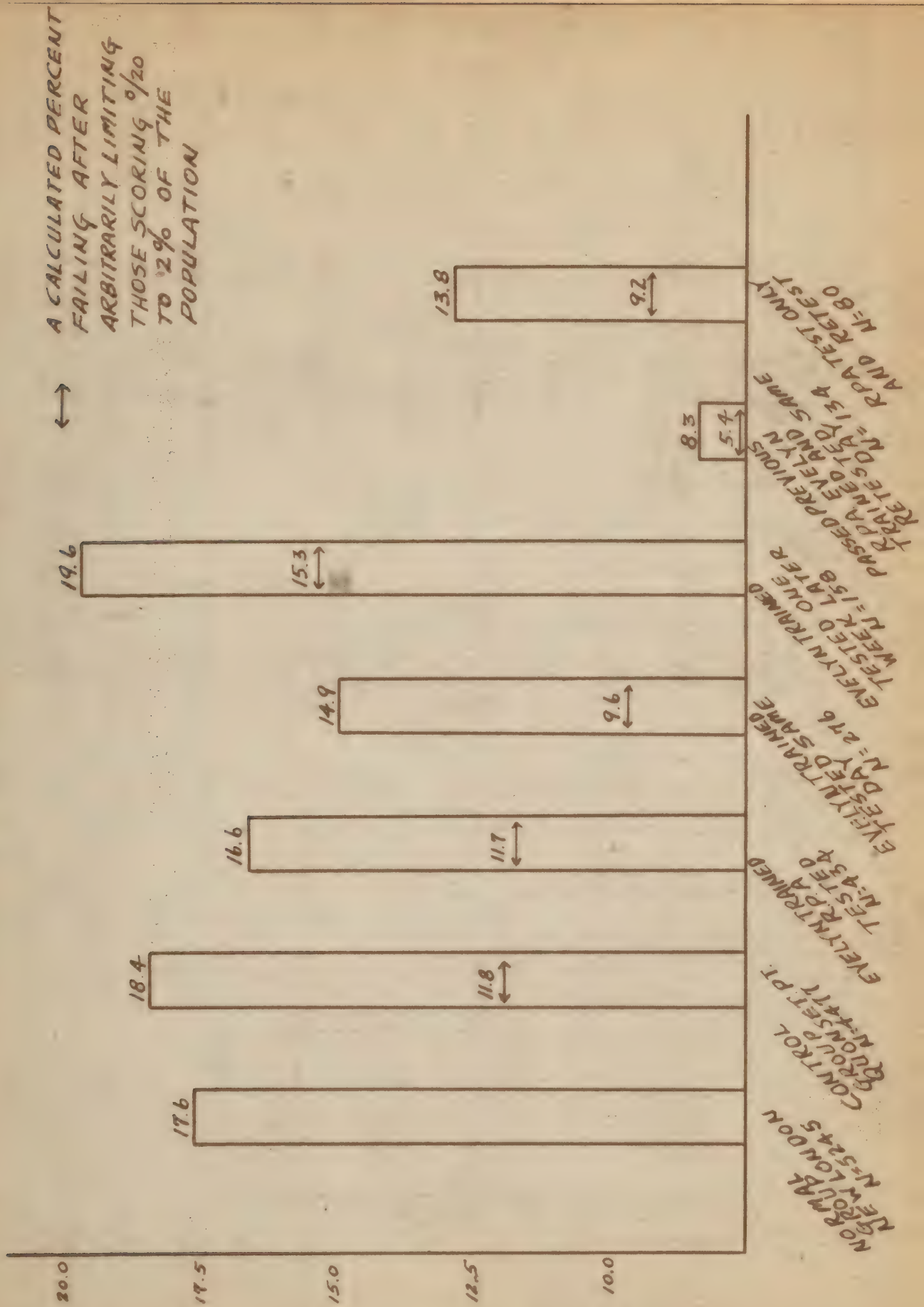DISTRIBUTION OF R.P.A. SCORES AT QUONSET POINT
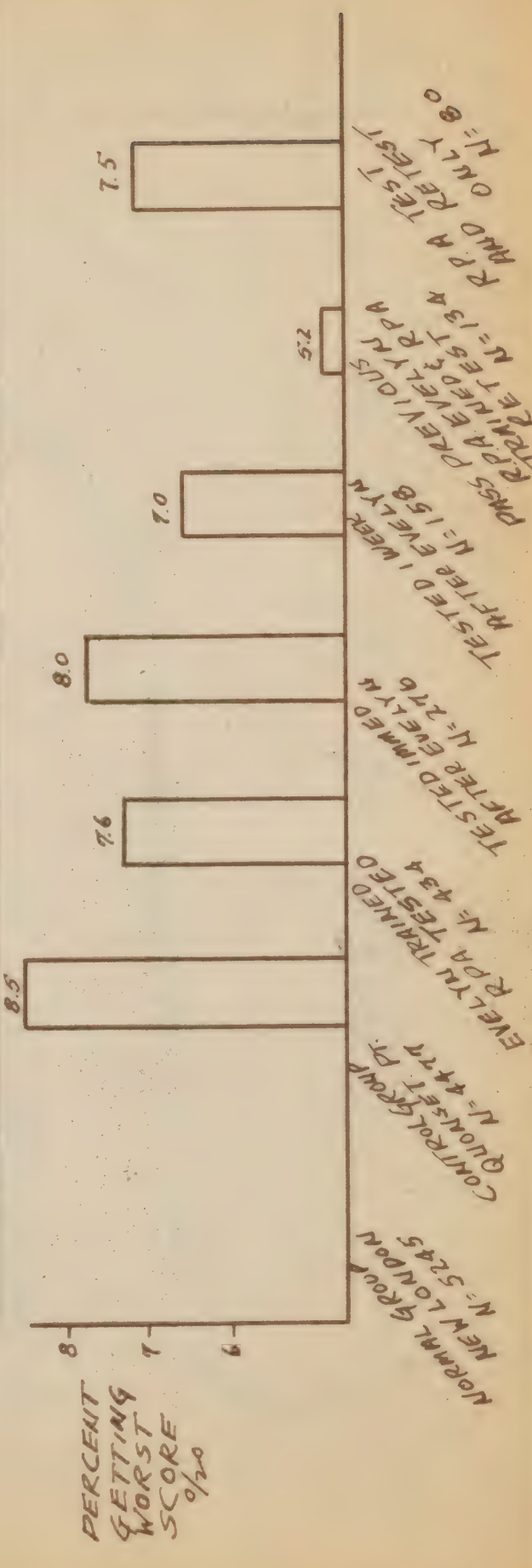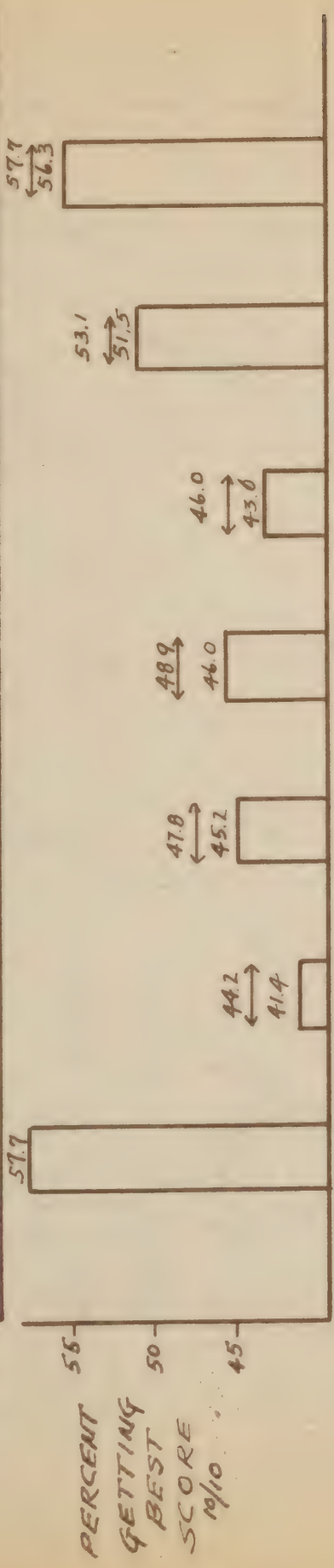
JUNE 1944 — FEB 1945 (N = 4471)

PERCENT

RPA SCORE

"FAIL"
(18.4%)

"PASS"

FIGURE 1

PERCENT FAILING R.P.A TEST

STANDARD AND TEST GROUPS

↕ A CALCULATED PERCENT
FAILING AFTER
ARBITRARILY LIMITING
THOSE SCORING %/20
TO '2% OF THE
POPULATION

PERCENT FAILING RPA TEST

CORRECTED INDICES SEE FIG 2

THIED
MEDIAN
SCORE

19.00
18.80
18.60
18.40
18.20

19.16
19.11
19.06
19.03
18.67
18.44
18.92
18.69
18.83
18.01
18.55
18.28

PERCENT
GETTING
BEST
SCORE %/10

55
50
45

57.1
56.3
53.1
51.5
46.0
43.6
48.2
46.0
47.8
45.2
44.2
41.4
51.7

PERCENT
GETTING
WORST
SCORE %/20

8
7
6

7.5
5.2
7.0
8.0
7.6
8.5

NORMAL LONDON NEW N=525
CONTROLS & TESTED PT QUOINSETTA N=4174
EVELYN TRAINED & TESTED 1st N=276
TESTED IMMED AFTER EVELYN N=2776
TESTED 1 WEEK AFTER EVELYN N=158
PASS EXPERIMENT N=13-4 PREVIOUS/PA
RPA RETEST N=0 ONLY 80

CUMULATIVE PERCENT CURVES